Aquatic Species Mapping in North Carolina Using Maxent

Mark Endries

U.S. Fish and Wildlife Service, Ecological Services Field Office, Asheville North Carolina

INTRODUCTION

The mission of the U.S. Fish and Wildlife Service (Service) is to work with others to conserve, protect, and enhance fish, wildlife, and plants and their habitats for the continuing benefit of the American people. The Service is the lead governmental agency involved in the recovery of federally endangered and threatened species in freshwater and terrestrial habitats. To meet its recovery and protection goals, the Service: (1) works with other federal agencies to minimize or eliminate impacts to fish, wildlife, and plants from projects they authorize, fund, or carry out; (2) supports the improvement of fish and wildlife habitat on private land through technical and financial assistance; and (3) provides scientific knowledge and analyses to help guide the conservation, development, and management of the Nation's fish and wildlife resources.

Freshwater ecosystems present unique management challenges due to their linear spatial orientation and their association with upland habitat variables. On broad scales, the movement of aquatic species within the stream environment is limited to upstream and downstream migration. The inability of aquatic species to circumnavigate man-made obstacles causes them to be particularly vulnerable to habitat fragmentation. Habitat fragmentation has a major influence on species distribution and complicates distribution mapping.

To better understand the spatial distributions of freshwater aquatic species in North Carolina, the Service created predictive habitat maps for 226 different aquatic species using geographic information systems (GIS) and maximum entropy (Maxent) modeling. These maps were derived by comparing known species occurrences with a suite of stream- or land-cover-derived environmental variables.

GIS provides an ideal tool for regional and statewide assessments of landscapes, the development and application of habitat models, and modeling of the potential distribution of species and habitats (Conner and Leopold 1998, Stoms et al. 1992). GIS is also an effective tool to assist in the resolution of land-use conflict and the management of natural resources (Brown et al. 1994). Given appropriate digital habitat and wildlife data, these data can be used to identify environmentally sensitive land, to allow GIS users to view their project in a landscape perspective, and to allow habitat quality and wildlife needs to be simulated as a function of proposed management (Conner and Leopold 1998).

Maxent is a machine learning technique that can be used to predict the geographic distribution of any spatial phenomena, including plants and animals. Its origins lie in statistical mechanics (Jaynes 1957). The Bayesian probability postulate of maximum entropy states that, subject to known constraints, the probability distribution which best represents the current state of knowledge is the one with largest entropy. Maxent compares a set of occurrences to a set of environmental variables of the same defined space to estimate a target probability distribution of maximum entropy. Unlike other techniques, such as logistic regression, Maxent does not require species absence data. Maxent can also produce valid output with a small set of observations; however, with larger sets of observations, models are typically more accurate.

METHODS

Study Area

The analysis extent was the State of North Carolina.

Source Data

Aquatic Species Occurrence Data

Aquatic species point-occurrence data was identified from six sources:

1. North Carolina Natural Heritage Program (NCNHP) Element Occurrence Dataset (NCNHP 2013).

2. North Carolina Museum of Natural Sciences, Research and Collections Section Dataset (North Carolina Museum of Natural Sciences 2013).

3. North Carolina Wildlife Resources Commission (NCWRC), Priority Species Monitoring Dataset (NCWRC 2013a).

4. NCWRC, Trout Distribution Dataset (NCWRC 2013b).

Only data that was collected from the year 2000 to the present was used in order to satisfy the requirement of Maxent for temporal correspondence between occurrence localities and environmental variables (Phillips et al. 2006). The land cover map that was used to create many of the environmental variables was based on satellite imagery from 2000.

From initial inspection, it was identified that many of the individual data points overlapped between datasets. To remove the potential for erroneous data duplication, the data was simplified to only identify the presence of a species by the individual stream segment. To do this, all species 0 occurrences that plotted within 100 meters of a digital stream segment were spatially identified by the center-point coordinate of the digital stream segment. Any point that did not plot within 100 meters of the digital stream dataset was not included in further analysis. The final step was to remove duplicate records where an individual stream segment was identified more than once by a unique species. The resultant aquatic species point-occurrence dataset identifies the center-point coordinate of each unique stream segment that a unique species had been sampled in from any of the six original point-occurrence datasets.

Digital Stream Dataset

Streams were represented by the 1:100,000-scale National Hydrography Dataset Plus Version 2 (NHDPlus) (Horizon Systems Corporation 2013). NHDPlus is a suite of application-ready geospatial products that builds upon, and extends, the capabilities of the original 1:100,000-scale National Hydrography Dataset (NHD) (U.S. Geological Survey [USGS] 2011a). NHDPlus integrates the NHD with the National Elevation Dataset (Fry et al. 2011) and Watershed Boundary Dataset (Natural Resources Conservation Service [NRCS] 2013). It adds a set of value-added attributes to enhance stream network navigation, an elevation-based catchment for each flowline (segment), cumulative drainage area characteristics, flow direction and accumulation, flowline slopes, and flowline volume and velocity estimates.

Land Cover Dataset

Land cover was represented by the 2006 National Land Cover Database (NLCD 2006) (Fry et al. 2011). NLCD 2006 is a 16-class land cover classification scheme that has been applied consistently across the conterminous United States at a spatial resolution of 30 meters. NLCD 2006 is based primarily on a decision-tree classification of circa 2006 Landsat satellite data.

Environmental Variables

Sixteen different environmental variables were used in the Maxent modeling. Six were derived directly from the digital stream dataset, seven were derived by summarizing land cover percentages by stream segment catchment, one was derived from a geologic map of North Carolina, one was derived from a map of river basins in North Carolina, and one was derived by estimating the percentage of land cover disturbance within 100 meters of each stream segment. A description of each variable is given below.

Digital Stream-Dataset-Derived Environmental Variables

Six environmental variables were derived from the digital stream dataset for each stream segment (Table 1).

Table 1. Description of Digital Stream-Dataset-Derived Environmental Variables

Variable	Description
Drainage Area	Cumulative drainage area in km^2.
Flow Rate	Mean annual flow in cubic feet per second computed using the unit runoff method.

Velocity	Mean annual velocity in feet per second computed using the Jobson method.
Strahler Stream Order	A method of classifying stream segments based on the number of tributaries upstream. A stream with no tributaries is considered a first-order stream. Stream order increases with the number of tributaries upstream.
Gradient	Slope of stream segment.
Sinuosity	A measure of deviation of a path between two points from the shortest possible path. The ratio of the actual path length divided by the shortest path length.
Temperature	Mean annual temperature in area upstream of the bottom of the flowline segment.
Precipitation	Mean annual precipitation in area upstream of the bottom of the flowline segment.

For environmental variables, Maxent requires ASCII raster grids. As a result, the vector digital stream dataset was converted into ASCII raster grids by buffering each stream segment by 100 meters and converting the buffer dataset into 30-meter ASCII raster grids using the various variable attributes. Each resultant ASCII raster's pixel value is the variable value of each stream segment within 100 meters of the stream segment. Areas outside of 100 meters of a stream segment are not given any value.

Land-Cover-Derived Environmental Variables

The land-cover-derived environmental variables summarize land cover percentages by NHDPlus stream segment catchment. The NHDPlus catchment layer identifies the extent of land that drains into each stream segment. The NLCD 2006 was reclassified six times into six general land cover grouping raster datasets (Table 2). Originally, ten generalized land cover groupings were created. Due to excessive correlation between the groupings, four of the original ten layers were removed from further analysis.

Table 2. Description of generalized land cover groupings

Land Cover Grouping	Description
Barren Land	Barren Land landcover class only.
Crop Land	Cultivated Crops landcover class only.
Developed	All developed land cover classes. Includes Developed Open Space, Developed Low Intensity, Developed Medium Intensity, Developed High Intensity
Forest Land	All forested land cover classes. Includes Deciduous Forest, Evergreen Forest, and Mixed Forest classes.
Pasture Land	Includes Herbaceous and Hay/Pasture landcover classes.
Shrub Land	Shrub/Scrub landcover class only.
Wetland	Includes Woody Wetlands and Emergent Herbaceous Wetlands classes.

To generate the catchment land cover percentage variable datasets, the "Tabulate Area" command in ArcGIS was performed to identify the amount of land cover grouping in each catchment. Next, the amount of land cover grouping was divided by the total catchment area; that value was multiplied by 100 to get the percentage and was then joined back to the original vector catchment dataset. The final step was to convert the catchment datasets to 30-meter ASCII raster datasets using the percent land cover grouping values. Each resultant ASCII raster pixel value is the variable percentage value for each stream segment catchment.

Percent Impervious Variable

Similar to the land-cover-derived variables, this variable identifies the percent of each catchment that is an impervious surface. The source for impervious surface data was the 2006 National Land Cover Dataset Percent Developed Imperviousness (USGS 2011d). The imperviousness dataset ranks the amount of imperviousness by pixel. The imperviousness dataset was reclassified by retaining all pixels with only a 20 percent or greater amount of imperviousness to represent areas impervious. To generate the catchment percent impervious variable dataset, a "Tabulate Area" command was performed in ArcGIS to identify the amount of impervious surface in each catchment. Next, the amount of impervious surface was divided by the total catchment area; that value was multiplied by 100 to get the percentage and was then joined back to the original vector catchment dataset. The final step was to convert the catchment dataset to a 30-meter ASCII raster datasets using the percent impervious surface grouping values. Each resultant ASCII raster pixel value is the catchment percent impervious value.

Geology Variable

This variable identifies the geology of the study area. The source of the data is a state geology map (1:500,000-scale) (North Carolina Geological Survey 2007). The original vector data was converted to a 30-meter ASCII raster.

River Basin Variable

This variable identifies the HUC-6 river basins in the study area (NRCS 2011). The original vector data was converted to a 30-meter ASCII raster.

Percent Riparian Disturbance Variable

This variable estimates the amount of disturbance within 100 meters surrounding the NHDPlus stream segments. The NLCD 2006 was reclassified to identify all disturbance categories as a single value. The disturbance land categories included the Developed land cover categories, Barren Land, Hay/Pasture, and Cultivated Crops. To identify the percent riparian disturbance by stream segment, a "Tabulate Area" command in ArcGIS was performed to identify the amount of riparian disturbance surface in each 100-meter buffer of the stream segments. Next, the amount of riparian disturbance was divided by the total river segment buffer area; that value was multiplied by 100 to get the percentage. This value was then joined back to the original vector stream segment buffer dataset. The final step was to convert the stream segment buffer dataset to a 30-meter ASCII raster dataset using the percent riparian disturbance grouping values. Each resultant ASCII raster pixel value is the stream segment riparian disturbance grouping values.

Model Creation

Maxent

Maxent software for species habitat modeling (version 3.3.3k) was used (Schapire 2013). Models were only generated for species with a minimum of 20 unique stream segments identified. Pearson et al. (2007) demonstrated that it was possible to accurately predict presences with as few as 5 to 12 occurrence records. Numerous other studies report models built with less than 25 records (Wisz et al. 2008, Papes and Gaubert 2007, Hernandez et al. 2006).

The specific settings used to run the Maxent models are as follows: A random test percentage of 20 was used. This withheld 20 percent of a species' occurrences and tested the validity and accuracy of the training model. Response curves and jackknife measures of variable importance were created to provide additional descriptive data and graphs to help interpret the Maxent model. The logistic output format was selected. Logistic output gives an estimate between 0 and 1. The cutoff value (to determine that typical presence localities have probability of presence, i.e., a stream segment is considered potential habitat for a species) used was the 10th percentile training presence. The 10th percentile training presence uses the suitability threshold associated with the presence record that occurs at the 10th percentile of presence records (Phillips 2011). This value excluded some of the outlier values included in the predictions however includes the majority of the occurrences in the model predictions.

Maxent creates a map of the logistic habitat predictions in ASCII format. The Maxent ASCII output was attributed to the NHD Plus V2 stream dataset multiplying each ASCII layer by 100 and attributing the values to the stream dataset. Storing the model output as a 0 to 100 value range resulted in significant file size reduction from the original ASCII. The final model raster datasets have a data range from 0 to 100, with 100 equaling 1 in the original ASCII output. Binary maps of species occurrence were created by reclassifying the integer raster datasets, where all values above the 10th percentile training presence logistic threshold (different for every species) were given a value of 1 and all other values were given a value of 0. A value of 1 indicates predicted species occurrence; a value of 0 indicates unsuitable habitat.

RESULTS

Maxent models for 247 different species were created--169 fish, 49 mussels, 27 crayfish, 1 amphibian, and 1 plant. The full list of species is given in Table 3. In general, Maxent did a good job in creating predictive habitat maps for all species. Training AUC values ranged from 0.8664 to 0.9987 with a mean of 0.9760, test AUC values ranged from 0.6853 to 0.998 with a mean of 0.9465. No species models needed to be discarded due to low AUC scores. The 10th percentile training presence values used to determine cutoff values in the models ranged from 0.1816 to 0.6295 with a mean of 0.3288.

Table 3. List of species with a Maxent model

Scientific Name	Common Name	Phylum	NatureServe Rank	Federal Status
Acantharchus pomotis	Mud Sunfish	Chordata	G5	
Alasmidonta heterodon	Dwarf Wedgemussel	Mollusca	G1G2	E
Alasmidonta raveneliana	Appalachian Elktoe	Mollusca	G1	E
Alasmidonta undulata	Triangle Floater	Mollusca	G4	
Alasmidonta varicosa	Brook Floater	Mollusca	G3	FSC
Alasmidonta viridis	Slippershell Mussel	Mollusca	G4G5	
Alosa sapidissima	American Shad	Chordata	G5	
Ambloplites cavifrons	Roanoke Bass	Chordata	G3	FSC
Ambloplites rupestris	Rock Bass	Chordata	G5	
Ameiurus brunneus	Snail Bullhead	Chordata	G4	
Ameiurus catus	White Catfish	Chordata	G5	
Ameiurus natalis	Yellow Bullhead	Chordata	G5	

Scientific Name	Common Name	Phylum	NatureServe Rank	Federal Status
Ameiurus nebulosus	Brown Bullhead	Chordata	G5	
Ameiurus platycephalus	Flat Bullhead	Chordata	G5	
Amia calva	Bowfin	Chordata	G5	
Anguilla rostrata	American Eel	Chordata	G4	FSC
Aphredoderus sayanus	Pirate Perch	Chordata	G5	
Cambarus asperimanus	Mitten Crayfish	Arthropoda	G4	
Cambarus bartonii	Appalachian Brook Crayfish	Arthropoda	G5T5	
Cambarus hiwasseensis	Hiwassee Crayfish	Arthropoda	G3G4	
Cambarus hobbsorum	Rocky River Crayfish	Arthropoda	G3G4	
Cambarus hystricosus	Sandhills Spiny Crayfish	Arthropoda	G2	
Cambarus johni	Carolina Foothills Crayfish	Arthropoda	G3	
Cambarus latimanus	Variable Crayfish	Arthropoda	G5	
Cambarus lenati	Broad River Crayfish	Arthropoda	G2	
Cambarus longulus	Atlantic Slope Crayfish	Arthropoda	G5	
Cambarus robustus	Big Water Crayfish	Arthropoda	G5	
Cambarus sp. C	A Crayfish	Arthropoda	G5	
Campeloma decisum	Pointed Campeloma	Mollusca	G5	
Campeloma limum	File Campeloma	Mollusca	G5	
Campostoma anomalum	Central Stoneroller	Chordata	G5	
Catostomus commersonii	White Sucker	Chordata	G5	
Centrarchus macropterus	Flier	Chordata	G5	
Chologaster cornuta	Swampfish	Chordata	G5	
Clinostomus funduloides	Rosyside Dace	Chordata	G5	FSC
Clinostomus sp. 1	Smoky Dace	Chordata	G5T3Q	
Cottus bairdi	Mottled Sculpin	Chordata	G5	
Cryptobranchus alleganiensis	Eastern Hellbender	Chordata	G3G4	FSC
Cyprinella analostana	Satinfin Shiner	Chordata	G5	
Cyprinella chloristia	Greenfin Shiner	Chordata	G4	
Cyprinella galactura	Whitetail Shiner	Chordata	G5	
Cyprinella labrosa	Thicklip Shiner	Chordata	G4	
Cyprinella lutrensis	Red Shiner	Chordata	G5	
Cyprinella nivea	Whitefin Shiner	Chordata	G4	
Cyprinella pyrrhomelas	Fieryblack Shiner	Chordata	G4	
Cyprinella spiloptera	Spotfin Shiner	Chordata	G5	
Cyprinella zanema	Santee Chub	Chordata	G4	
Cyprinus carpio	Common Carp	Chordata	G5	
Dorosoma cepedianum	American Gizzard Shad	Chordata	G5	
Elassoma zonatum	Banded Pygmy Sunfish	Chordata	G5	
Elimia catenaria	Gravel Elimia	Mollusca	G4	
Elimia proxima	Sprite Elimia	Mollusca	G5	

Scientific Name	Common Name	Phylum	NatureServe Rank	Federal Status
Elimia virginica	Piedmont Elimia	Mollusca	G5	
Elliptio angustata	Carolina Lance	Mollusca	G4	
Elliptio cistellaeformis	Box Spike	Mollusca	G4	
Elliptio complanata	Eastern Elliptio	Mollusca	G5	
Elliptio congaraea	Carolina Slabshell	Mollusca	G3	
Elliptio dilatata	Spike	Mollusca	G5	
Elliptio fisheriana	Northern Lance	Mollusca	G4	
Elliptio icterina	Variable Spike	Mollusca	G5Q	
Elliptio lanceolata	Yellow Lance	Mollusca	G2G3	FSC
Elliptio producta	Atlantic Spike	Mollusca	G3Q	
Elliptio roanokensis	Roanoke Slabshell	Mollusca	G2	
Elliptio viridula	Green Lance	Mollusca	G5	
Enneacanthus gloriosus	Bluespotted Sunfish	Chordata	G5	
Enneacanthus obesus	Banded Sunfish	Chordata	G5	
Erimonax monachus	Spotfin Chub	Chordata	G2	T
Erimystax insignis eristigma	Mountain Blotched Chub	Chordata	G4TNR	
Erimyzon oblongus	Creek Chubsucker	Chordata	G5	
Esox americanus	Redfin Pickerel	Chordata	G5	
Esox niger	Chain Pickerel	Chordata	G5	
Etheostoma blennioides	Greenside Darter	Chordata	G5	
Etheostoma brevispinum	Carolina Fantail Darter	Chordata	G4	
Etheostoma chlorobranchium	Greenfin Darter	Chordata	G4	FSC
Etheostoma collis	Carolina Darter	Chordata	G3	FSC
Etheostoma flabellare	Fantail Darter	Chordata	G5	
Etheostoma fusiforme	Swamp Darter	Chordata	G5	
Etheostoma gutselli	Tuckasegee Darter	Chordata	G4	
Etheostoma kanawhae	Kanawha Darter	Chordata	G4	
Etheostoma mariae	Pinewoods Darter	Chordata	G3	FSC
Etheostoma nigrum	Johnny Darter	Chordata	G5	
Etheostoma olmstedi	Tessellated Darter	Chordata	G5	
Etheostoma podostemone	Riverweed Darter	Chordata	G4	
Etheostoma rufilineatum	Redline Darter	Chordata	G5	
Etheostoma serrifer	Sawcheek Darter	Chordata	G5	
Etheostoma swannanoa	Swannanoa Darter	Chordata	G4	
Etheostoma thalassinum	Seagreen Darter	Chordata	G4	
Etheostoma vitreum	Glassy Darter	Chordata	G4G5	
Etheostoma vulneratum	Wounded Darter	Chordata	G3	FSC
Etheostoma zonale	Banded Darter	Chordata	G5	
Exoglossum laurae	Tonguetied Minnow	Chordata	G4	
Fundulus rathbuni	Speckled Killifish	Chordata	G4	

Scientific Name	Common Name	Phylum	NatureServe Rank	Federal Status
Fusconaia masoni	Atlantic Pigtoe	Mollusca	G2	FSC
Gambusia holbrooki	Eastern Mosquitofish	Chordata	G5	
Helisoma anceps	Two-Ridge Rams-Horn	Mollusca	G5	
Hybognathus regius	Eastern Silvery Minnow	Chordata	G5	
Hybopsis amblops	Bigeye Chub	Chordata	G5	
Hybopsis hypsinotus	Highback Chub	Chordata	G4	
Hypentelium nigricans	Northern Hog Sucker	Chordata	G5	
Hypentelium roanokense	Roanoke Hog Sucker	Chordata	G4	
Ichthyomyzon greeleyi	Mountain Brook Lamprey	Chordata	G3G4	
Ictalurus punctatus	Channel Catfish	Chordata	G5	
Lampsilis cariosa	Yellow Lampmussel	Mollusca	G3G4	FSC
Lampsilis fasciola	Wavy-Rayed Lampmussel	Mollusca	G5	
Lampsilis radiata	Eastern Lampmussel	Mollusca	G5	
Lampsilis sp. 2	Chameleon Lampmussel	Mollusca	G1	
Lasmigona subviridis	Green Floater	Mollusca	G3	FSC
Lepisosteus osseus	Longnose Gar	Chordata	G5	
Lepomis auritus	Redbreast Sunfish	Chordata	G5	
Lepomis cyanellus	Green Sunfish	Chordata	G5	
Lepomis gibbosus	Kiver / Pumpkinseed	Chordata	G5	
Lepomis gulosus	Warmouth	Chordata	G5	
Lepomis macrochirus	Bluegill	Chordata	G5	
Lepomis marginatus	Dollar Sunfish	Chordata	G5	
Lepomis microlophus	Redear Sunfish	Chordata	G5	
Lepomis punctatus	Spotted Sunfish	Chordata	G5	
Leptodea ochracea	Tidewater Mucket	Mollusca	G3G4	
Leptoxis carinata	Crested Mudalia	Mollusca	G5	
Leptoxis dilatata	Seep Mudalia	Mollusca	G3	
Luxilus albeolus	White Shiner	Chordata	G5	
Luxilus cerasinus	Crescent Shiner	Chordata	G4	
Luxilus chrysocephalus	Striped Shiner	Chordata	G5	
Luxilus coccogenis	Warpaint Shiner	Chordata	G5	
Lythrurus ardens	Rosefin Shiner	Chordata	G5	
Lythrurus matutinus	Pinewoods Shiner	Chordata	G3	FSC
Micromenetus dilatatus	Bugle Sprite	Mollusca	G5	
Micropterus dolomieu	Smallmouth Bass	Chordata	G5	
Micropterus punctulatus	Spotted Bass	Chordata	G5	
Micropterus salmoides	Largemouth Bass	Chordata	G5	
Minytrema melanops	Spotted Sucker	Chordata	G5	
Morone americana	White Perch	Chordata	G5	
Moxostoma anisurum	Silver Redhorse	Chordata	G5	

Scientific Name	Common Name	Phylum	NatureServe Rank	Federal Status
Moxostoma ariommum	Bigeye Jumprock	Chordata	G4	
Moxostoma breviceps	Hookfin Redhorse	Chordata	G5	
Moxostoma carinatum	River Redhorse	Chordata	G4	
Moxostoma cervinum	Blacktip Jumprock	Chordata	G4	
Moxostoma collapsum	Piedmont Redhorse	Chordata	G5	
Moxostoma duquesnei	Black Redhorse	Chordata	G5	
Moxostoma erythrurum	Golden Redhorse	Chordata	G5	
Moxostoma macrolepidotum	Shorthead Redhorse	Chordata	G5	
Moxostoma pappillosum	Slender Redhorse	Chordata	G4	
Moxostoma rupiscartes	Striped Jumprock	Chordata	G4	
Moxostoma sp. 2	Sicklefin Redhorse	Chordata	G2Q	C
Moxostoma sp. 3	Carolina Redhorse	Chordata	G1G2Q	
Moxostoma sp. nov.	Brassy Jumprock	Chordata	G4	
Necturus lewisi	Neuse River Waterdog	Chordata	G3	
Nocomis leptocephalus	Bluehead Chub	Chordata	G5	
Nocomis micropogon	River Chub	Chordata	G5	
Nocomis platyrhynchus	Bigmouth Chub	Chordata	G4Q	
Nocomis raneyi	Bull Chub	Chordata	G4	
Notemigonus crysoleucas	Golden Shiner	Chordata	G5	
Notropis alborus	Whitemouth Shiner	Chordata	G4	
Notropis altipinnis	Highfin Shiner	Chordata	G5	
Notropis amoenus	Comely Shiner	Chordata	G5	
Notropis chalybaeus	Ironcolor Shiner	Chordata	G4	
Notropis chiliticus	Redlip Shiner	Chordata	G4	
Notropis chlorocephalus	Greenhead Shiner	Chordata	G4	
Notropis cummingsae	Dusky Shiner	Chordata	G5	
Notropis hudsonius	Spottail Shiner	Chordata	G5	
Notropis leuciodus	Tennessee Shiner	Chordata	G5	
Notropis lutipinnis	Yellowfin Shiner	Chordata	G4Q	
Notropis mekistocholas	Cape Fear Shiner	Chordata	G1	E
Notropis micropteryx	Highland Shiner	Chordata	G5	
Notropis petersoni	Coastal Shiner	Chordata	G5	
Notropis photogenis	Silver Shiner	Chordata	G5	
Notropis procne	Swallowtail Shiner	Chordata	G5	
Notropis rubricroceus	Saffron Shiner	Chordata	G4G5	
Notropis scabriceps	New River Shiner	Chordata	G4	
Notropis scepticus	Sandbar Shiner	Chordata	G4	
Notropis sp. 1	Rosyface Shiner - Upper New River	Chordata	G5	
Notropis sp. cf. chlorocephalus	Piedmont Shiner	Chordata	G5	
Notropis spectrunculus	Mirror Shiner	Chordata	G4	

Scientific Name	Common Name	Phylum	NatureServe Rank	Federal Status
Notropis telescopus	Telescope Shiner	Chordata	G5	
Notropis volucellus	Mimic Shiner	Chordata	G5	
Noturus furiosus	Carolina Madtom	Chordata	G2	FSC
Noturus gyrinus	Tadpole Madtom	Chordata	G5	
Noturus insignis	Margined Madtom	Chordata	G5	
Oncorhynchus mykiss	Rainbow Trout	Chordata	G5	
Orconectes carolinensis	North Carolina Spiny Crayfish	Arthropoda	G3	
Perca flavescens	Yellow Perch	Chordata	G5	
Percina aurantiaca	Tangerine Darter	Chordata	G4	
Percina crassa	Piedmont Darter	Chordata	G4	
Percina evides	Gilt Darter	Chordata	G4	
Percina gymnocephala	Appalachia Darter	Chordata	G4	
Percina nevisense	Chainback Darter	Chordata	G4G5	
Percina peltata	Shield Darter	Chordata	G5	
Percina roanoka	Roanoke Darter	Chordata	G4	
Percina squamata	Olive Darter	Chordata	G3	FSC
Phenacobius crassilabrum	Fatlips Minnow	Chordata	G3G4	
Phenacobius teretulus	Kanawha Minnow	Chordata	G3G4	FSC
Phoxinus oreas	Mountain Redbelly Dace	Chordata	G5	
Pimephales promelas	Fathead Minnow	Chordata	G5	
Planorbella trivolvis	Marsh Rams-Horn	Mollusca	G5	
Pleurobema collina	James River Spinymussel	Mollusca	G1	E
Pomoxis nigromaculatus	Black Crappie	Chordata	G5	
Procambarus acutus	White River Crawfish	Arthropoda	G5	
Psuedosuccinea columella	American Ribbed Fluke Snail	Mollusca	G5	
Pyganodon cataracta	Eastern Floater	Mollusca	G5	
Pylodictis olivaris	Flathead Catfish	Chordata	G5	
Rhinichthys atratulus	Blacknose Dace	Chordata	G5	
Rhinichthys cataractae	Longnose Dace	Chordata	G5	
Rhinichthys obtusus	Western Blacknose Dace	Chordata	G5	
Salmo trutta	Brown Trout	Chordata	G5	
Salvelinus fontinalis	Brook Trout	Chordata	G5	
Semotilus atromaculatus	Creek Chub	Chordata	G5	
Semotilus lumbee	Sandhills Chub	Chordata	G3	FSC
Strophitus undulatus	Creeper / Squawfoot	Mollusca	G5	
Trinectes maculatus	Hogchoker	Chordata	G5	
Umbra pygmaea	Eastern Mudminnow	Chordata	G5	
Uniomerus carolinianus	Florida Pondhorn	Mollusca	G4	
Utterbackia imbecillis	Paper Pondshell	Mollusca	G5	
Villosa constricta	Notched Rainbow	Mollusca	G3	

Scientific Name	Common Name	Phylum	NatureServe Rank	Federal Status
Villosa delumbis	Eastern Creekshell	Mollusca	G4	
Villosa iris	Rainbow	Mollusca	G5Q	
Villosa vaughaniana	Carolina Creekshell	Mollusca	G2	FSC
Pegias fabula	Littlewing Pearlymussel	Mollusca	G1	E
Menidia extensa	Waccamaw Silverside	Chordata	G1	T
Lasmigona decorata	Carolina Heelsplitter	Mollusca	G1	E
Acipenser brevirostrum	Shortnose Sturgeon	Chordata	G3	E
Cambarus acuminatus	Acuminate Crayfish	Arthropoda	G4Q	
Cambarus chasmodactylus	New River Crayfish	Arthropoda	G4	FSC
Cambarus diogenes	Devil Crawfish	Arthropoda	G5	
Cambarus dubius	Upland Burrowing Crayfish	Arthropoda	G5	
Cambarus howardi	Chattahoochee Crayfish	Arthropoda	G3Q	
Cambarus longirostris	Longnose Crayfish	Arthropoda	G5Q	
Cambarus reburrus	French Broad Crayfish	Arthropoda	G3	FSC

In an attempt to create a statewide prioritization of all streams in North Carolina, all Maxent model predictions were summarized based on NatureServe's species of global rank (NatureServe 2011). NatureServe's global rank provides an assessment of the condition of a species across its entire range and includes an estimate of extinction risk. The statewide prioritization ranks streams in North Carolina based on species diversity and global rank (see Figure 1). To create the statewide prioritization, for each NHD stream segment the global rank for each species predicted to be present was identified, and the sum of these global ranks was calculated using the global rank value scheme given in Table 4. An exponential scale was used to create the calculation value as a way to give increasing weight to the greater levels of imperilment and species extinction risk as the global rank increases from G5 to G1.

Table 4. Global rank classification value scheme.

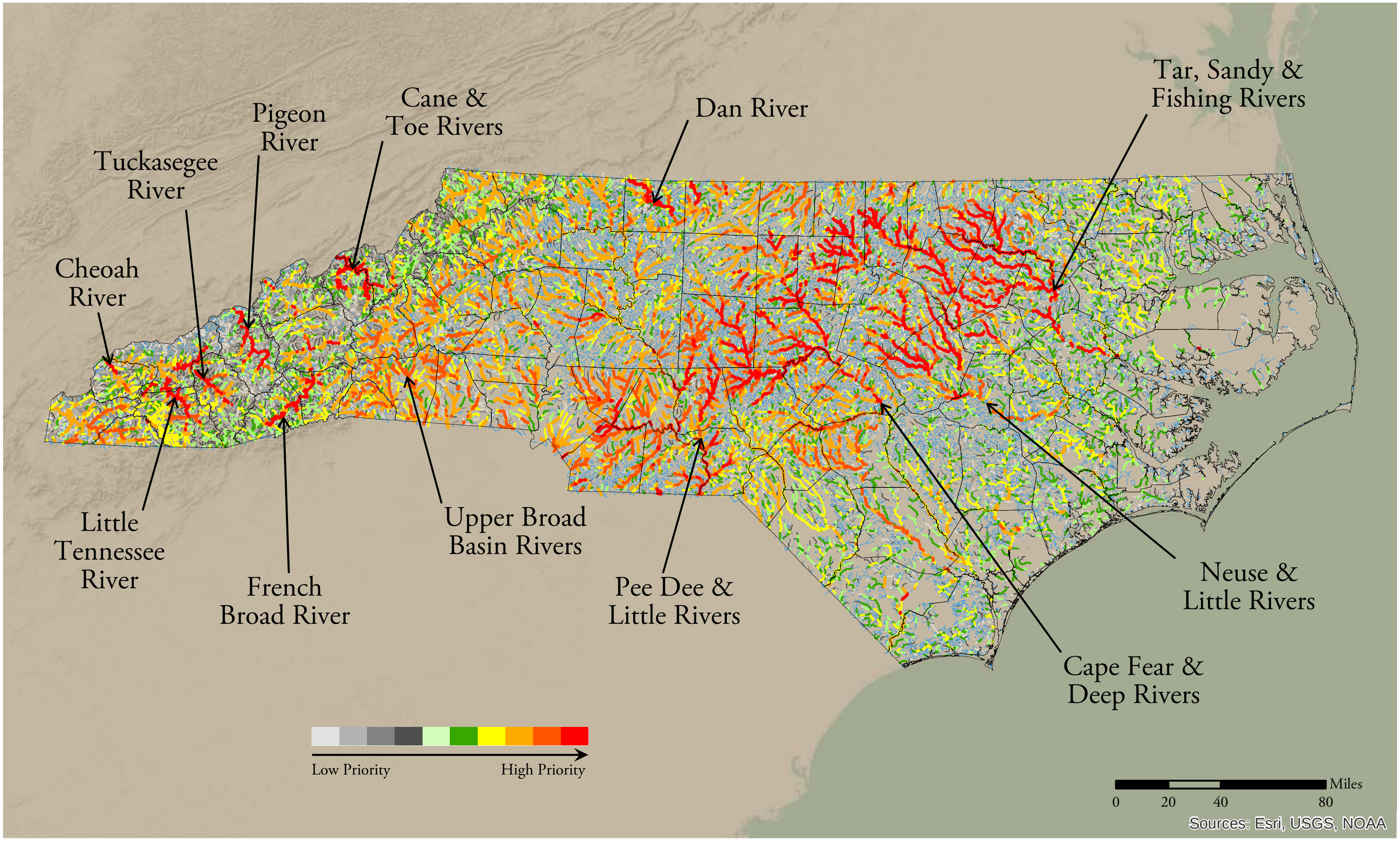
Calculation Value	Global Rank
0	No species
1	G5 species
2	G4G5 species
4	G4 species
8	G3G4 species
16	G3 species
32	G2G3 species
64	G2 species
128	G1G2 species
256	G1 species

The resultant calculation was classified using a 10-class scheme given in Table 5 to produce the final statewide prioritization layer. An exponential scale was again used to generate a 10-class classification scheme.

Table 5. Global rank classification value scheme.

Value	Calculation Sum
0	0
1	1
2	2
3	3 – 4
4	5 – 8
5	9 – 16
6	17 – 32
7	33 – 64
8	65 – 128
9	129 – 256
10	Greater than 257

Figure 1. Statewide prioritization of streams based on species global ranks and diversity. Areas with pointers indicate highest priority areas in North Carolina.



DISCUSSION

Fifty-four percent of all streams were identified as predicted habitat by at least one species. Just 19 percent of all streams were identified as predicted habitat that includes at least one G3, G2, or G1 species. It is surprising that roughly half of all the streams in western North Carolina were not predicted to be potential habitat for any species included in the analysis. However, many of these unpredicted streams were headwater segments. It is likely, due to the detail of the NHD stream dataset, that many of these headwater segments are intermittent and therefore do not maintain persistent populations of aquatic organisms.

Table 6. Counts of variable contributions to the gain in Maxent model prediction power.

Variable	Top Contributor	≥10-Percent Contribution
Barren Land	0	0
Crop Land	0	7
Drainage Area	114	154
Flow Rate	27	118
Forest Land	0	2
Geology	25	122
Gradient	0	12
Impervious Surface	0	2
Pasture Land	0	7
Percent Disturbed	0	5
Precipitation	9	49
Shrub Land	0	1
Sinuosity	0	1
Strahler Stream Order	3	62
Temperature	49	109
Velocity	8	21
Wetland	12	51

The drainage area variable was both the variable that most often was the top predictive variable (n = 114) and the variable that most often contributed 10 percent or more to the gain in prediction power of a model (n = 154) (Table 6). This variable is related to the River Continuum Concept (RCC) (Vannote et al. 1980). The basic concept of RCC is that streams exhibit a continuous gradient of physical characteristics and available energy that correspond to a resulting gradient in species composition and richness. The distribution of a species within the continuum is based on that species niche and corresponds to associated community type. Within similar geographic regions, these stream-derived variables can be expected to predict community type and should be good predictors of species distribution. Additional variables that are also related to the RCC are flow rate, Strahler stream order, and velocity. The flow rate variable is also a common top (n = 27) and

10 percent or more variable predictor (n = 118), Strahler stream order and velocity less so, but they both were repeatedly found to be both top and 10 percent or more variable predictors.

The temperature and precipitation variables are new to the NHDPlus dataset and both proved to be valuable in model predictions. Temperature was the second top contributing variable (n = 49) and a common variable that contributed at least 10 percent game in the prediction power of a model (n = 109). Temperature has long been recognized as an important environmental factor in aquatic ecosystems in regard to its pivotal role over biological (development, growth and reproduction), chemical, and physical properties. Aquatic organisms all have a preferred temperature range and cold-blooded species must maintain a specific internal temperature or inhabit environments within a temperature range. Precipitation regimes, including the volume, magnitude, timing, duration, frequency, and variation of precipitation events have broad effects on ecosystem productivity, habitat structure, and ultimately on resident fish, invertebrate, and algae communities.

The geology variable was the fourth most top contributing variable (n = 25), and the second most common variable to contribute at least 10 percent to the gain in prediction power of a model (n = 122). Geology plays an important role in species distribution in two ways. Geology is related to the physical characteristics of a stream by its influence on stream form, bedload transport characteristics, and substrate composition, factors largely determining habitat type. Geology is also influential in determining the chemical characteristics of the water and the substrate, factors that play an important role in the distribution of some sensitive species. Due to its broad influence on multiple factors that affect habitat suitability, geology tends to be a generic predictor of community type, and community type has an influence on species distribution.

The only land-cover-derived variable that only was infrequently the top predictive variable in any of the models and contributed at least 10 percent to the gain in prediction power of a model was the percent wetland variable. The importance of wetlands to aquatic systems cannot be overstated. Wetlands help to stabilize river water levels, filter out sediment to purify water, and release nutrients.

It was hoped that the other land-cover-derived variables would identify areas where alterations to the land cover, or natural land cover compositions would have significant effects on aquatic species distributions, but this was largely not the case. It is possible that deficiency was due to errors in the land cover, errors associated with temporal disagreements between occurrences and the land cover map, or that aquatic species are impacted by land cover alterations at a scale different from that of a catchment. Also, the mapping efforts did not address the cumulative effects of land cover alterations within a river network.

Models of ecological processes are simplifications of staggeringly complex systems. As a result, ecological models have limitations in what they address and how they should be applied. The Maxent analysis was limited to coarse-scale variables at the segment and catchment scale and does not address stream microhabitat features such as substrate, bank stability, localized alterations, acute disturbances, etc., within a stream segment. These microhabitat features affect the spatial distribution of aquatic species yet are not addressed in the modeling efforts. For mussel distributions, the modeling does not take into consideration the distribution of host fishes. Also not included in the models are water quality variables such as point- and nonpoint-source pollution, water chemistry, siltation, and temperature. Finally, despite using four different datasets on aquatic species distributions, not all species occurrences are known.

The author feels that the maps created provide an excellent coarse-scale look at the potential stream suitability of many aquatic species present in North Carolina. Prior to any land-management action or decision, direct consultation with species and habitat experts and detailed stream surveys to verify the presence of species or appropriate habitat are needed. The data should serve only as a guide. It is hoped that the mapping efforts can help prioritize stream systems and help to educate people on the spatial distributions and conservation needs of aquatic species and habitats in North Carolina.

REFERENCES

Brown, S., H. Schreier, W. A. Thompson, and I. Vertinsky. 1994. Linking Multiple Accounts with GIS as Decision Support System to Resolve Forestry/Wildlife Conflicts. Journal of Environmental Management 42:349-364.

Conner, L. M., and B. D. Leopold. 1998. A Multivariate Habitat Model for Female Bobcats: A GIS Approach. Proceedings of the Annual Conference of the Southeastern Association of Fish and Wildlife Agencies 52:232-243.

Fry, J., Xian, G., Jin, S., Dewitz, J., Homer, C., Yang, L., Barnes, C., Herold, N., and Wickham, J., 2011.Completion of the 2006 National Land Cover Database for the Conterminous United States, *PE&RS*, Vol. 77(9):858-864.

Hernandez, P. A., C. H. Graham, L. L. Master, and D. L. Albert. 2006. The Effect of Sample Size and Species Characteristics on Performance of Different Species Distribution Modeling Methods. Ecography 29:773-785.

Horizon Systems Corporation. 2013. National Hydrography Dataset Plus Version 2. Available at <http://www.horizon-systems.com/nhdplus/NHDPlusV2_home.php>.

Jaynes, E. T. 1957. Information Theory and Statistical Mechanics. The Physical Review, 106(4): 620-630.

Natural Resources Conservation Service. 2011. Watershed Boundary Dataset. Available at http://www.ncgc.nrcs.usda.gov/products/datasets/watershed/.

NatureServe. 2013. NatureServe Conservation Status. Available at <http://www.natureserve.org/explorer/ranking.htm>.

North Carolina Geological Survey (NCGS). 2007. North Carolina Geology: NCGS, Raleigh, North Carolina. Available at <http://www.nconemap.com>.

North Carolina Natural Heritage Program. 2013. Heritage Data. Available at <http://www.ncnhp.org/Pages/heritagedata.html>.

North Carolina Museum of Natural Sciences. 2013. Research and Collections. Available at <http://naturalsciences.org/research-collections>.

North Carolina Wildlife Resources Commission. 2013a. Wildlife Species and Conservation. Available at <https://www.ncwildlife.org/give/index.htm>.

------. 2013b. Coldwater Hatcheries: Marion Hatchery. Available at <https://www.ncwildlife.org/Fishing/Fish_Coolwater_Hatcheries.htm#marion>.

Papes, M., and P. Gaubert. 2007. Modeling Ecological Niches from Low Numbers of Occurrences: Assessment of the Conservation Status of Poorly Known Vivverids (Mammalia, Carnivora) Across Two Continents. Diversity and Distributions 13:890-902.

Pearson, R. G., C. J. Raxworthy, M., Nakamura, and A. T. Pearson. 2007. Predicting Species Distributions from Small Numbers of Occurrence Records: A Test Case Using Cryptic Geckos in Madagascar. Journal of Biogeography 34:102-117.

Phillips, S. 2011. A Brief Tutorial on Maxent. Available at <http://www.cs.princeton.edu/~schapire/maxent/tutorial/tutorial.doc>.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum Entropy Modeling of Species Geographic Distributions. Ecological Modeling 190:231-259.

Schapire, R. E. 2013. Maxent software for species habitat modeling. Available at <http://www.cs.princeton.edu/~schapire/maxent/>.

Stoms, D. M., F. W. Davis, and C. B. Cogan. 1992. Sensitivity of Wildlife Habitat Models to Uncertainties in GIS Data. Photogrammetric Engineering and Remote Sensing 6:843-850.

U.S. Geological Survey. 2011a. National Hydrography Dataset. Available at <http://nhd.usgs.gov/>.

------. 2011b. National Elevation Dataset. Available at <http://ned.usgs.gov/>.

Vannote, R. L., G. W. Minshall, K. W. Cummins, J. R. Sedell, and C. E. Cushing. 1980. The river continuum concept. Canadian Journal of Fisheries and Aquatic Sciences 37:130-137.

Wisz, M. S., R. J. Hijmans, J. Li, A. T. Peterson, C. H. Graham, A. Guisan, and NCEAS Predicting Species Distributions Working Group. 2008. Effects of Sample Size on the Performance of Species Distribution Models. Diversity and Distributions 14:763-773.